

WHAT IS CLAIMED IS:

1 1. A method for detecting electronic text communication distributed
2 in bulk, the method comprising steps of:
3 receiving a first electronic text communication;
4 processing the first electronic text communication with an algorithm to
5 produce a first fingerprint;
6 beginning a time period for the first electronic text communication;
7 receiving a second electronic text communications;
8 processing the second electronic text communications with the algorithm
9 to produce a second fingerprint;
10 comparing the first fingerprint to the second fingerprint to determine if the
11 first electronic text communication is similar to the second electronic text
12 communication;
13 updating a count for the first electronic text communication based upon the
14 comparing step; and
15 determining if the count during the time period reaches a first threshold.

1 2. The method for detecting electronic text communication distributed
2 in bulk as recited in claim 1, further comprising a step of filtering subsequent electronic
3 text communications similar to the first electronic text communication.

1 3. The method for detecting electronic text communication distributed
2 in bulk as recited in claim 1, wherein the first listed processing step comprises a step of
3 calculating a histogram where counts are determined for words in the first electronic text
4 communication.

1 4. The method for detecting electronic text communication distributed
2 in bulk as recited in claim 1, further comprising steps of:
3 determining if a character count of the first electronic text communication
4 exceeds a second threshold; and
5 choosing a fingerprint algorithm based upon the step of determining if the
6 character count of the first electronic text communication exceeds the second threshold.

1 5. The method for detecting electronic text communication distributed
2 in bulk as recited in claim 1, wherein a match is determined from the comparing step even
3 if the first fingerprint and the second fingerprint differ by a percentage.

1 6. The method for detecting electronic text communication distributed
2 in bulk as recited in claim 1, further comprising steps of:
3 determining network addresses for the first and second electronic text
4 communications; and
5 modifying the first threshold based upon the step of determining network
6 addresses.

1 7. A method for detecting electronic text communication distributed
2 in bulk, the method comprising steps of:
3 receiving an electronic text communication;
4 processing the electronic text communication with an algorithm to produce
5 a fingerprint;
6 beginning a time period associated with the electronic text communication;
7 receiving a plurality of electronic text communications;
8 processing the plurality electronic text communications with the algorithm
9 to produce a plurality of fingerprints;
10 comparing the plurality of fingerprints to the fingerprint in order to
11 determine how many of the plurality of electronic text communications are similar to the
12 electronic text communication;
13 counting an amount of the plurality of electronic text communications that
14 are similar to the electronic text communication; and
15 determining if the amount during the time period reaches a first threshold.

1 8. The method for detecting electronic text communication distributed
2 in bulk as recited in claim 7, further comprising a step of filtering subsequent electronic
3 text communications similar to the electronic text communication.

1 9. The method for detecting electronic text communication distributed
2 in bulk as recited in claim 7, wherein the first listed processing step comprises a step of
3 calculating a histogram where counts are determined for words in the electronic text
4 communication.

1 10. The method for detecting electronic text communication distributed
2 in bulk as recited in claim 7, further comprising steps of:

3 determining if a character count of the electronic text communication
4 exceeds a second threshold; and

5 choosing a fingerprint algorithm based upon the step of determining if the
6 character count of the electronic text communication exceeds the second threshold.

1 11. The method for detecting electronic text communication distributed
2 in bulk as recited in claim 7, wherein the electronic text communication is chosen from a
3 group consisting of a chat room comment, an instant message, a newsgroup posting, an
4 electronic forum posting, a message board posting, and a classified advertisement.

1 12. The method for detecting electronic text communication distributed
2 in bulk as recited in claim 7, further comprising steps of:

3 determining network addresses for the electronic text communication and
4 each of the subset; and

5 modifying the first threshold based upon the step of determining network
6 addresses.

1 13. A method for blocking electronic text communication distributed in
2 bulk, the method comprising steps of:

3 receiving an electronic text communication;

4 generating a fingerprint indicative of the electronic text communication;

5 beginning a time period in relation to the first listed receiving step;

6 receiving a plurality of electronic text communications;

7 generating a plurality of fingerprints corresponding to the plurality of
8 electronic text communications;

9 determining a subset of the plurality of electronic text communications that
10 are similar to the electronic text communication;

11 counting a size of the subset;

12 determining if the size during the time period reaches a first threshold; and

13 filtering subsequent electronic text communications similar to the
14 electronic text communication.

1 14. The method for blocking electronic text communication distributed
2 in bulk as recited in claim 13, wherein the first listed generating step comprises a step of
3 calculating a histogram where counts are determined for words in the electronic text
4 communication.

1 15. The method for blocking electronic text communication distributed
2 in bulk as recited in claim 13, further comprising a step of removing non-textual
3 information from the electronic text communication.

1 16. The method for blocking electronic text communication distributed
2 in bulk as recited in claim 13, further comprising a step of determining if a character
3 count of the electronic text communication exceeds a second threshold.

1 17. The method for blocking electronic text communication distributed
2 in bulk as recited in claim 16, further comprising a step of choosing a fingerprint
3 algorithm based upon the step of determining if the character count of the electronic text
4 communication exceeds the second threshold.

1 18. The method for blocking electronic text communication distributed
2 in bulk as recited in claim 13, wherein the electronic text communication is chosen from a
3 group consisting of a chat room comment, an instant message, a newsgroup posting, an
4 electronic forum posting, a message board posting, and a classified advertisement.

1 19. The method for blocking electronic text communication distributed
2 in bulk as recited in claim 13, further comprising a step of removing everything from the
3 electronic text communication except a message body.

1 20. The method for blocking electronic text communication distributed
2 in bulk as recited in claim 13, further comprising steps of:
3 determining network addresses for the electronic text communication and
4 each of the subset; and
5 modifying the first threshold based upon the step of determining network
6 addresses.